2010

Evolution of an MHC Class La Gene Fragment in Four North American *Morone* Species

J. X. Liu

Bert Ely
University of South Carolina - Columbia, ely@sc.edu

Publication Info

Postprint version. Published in *Journal of Fish Biology*, Volume 76, 2010, pages 1984-1994.
This is the peer reviewed version of the following article: Evolution of an MHC Class La Gene Fragment in Four North American Morone Species, which has been published in final form at https://doi.org/10.1111/j.1095-8649.2010.02588.x. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions."

Evolution of an MHC class Ia gene fragment in four North American *Morone* species

Jin-Xian Liu, Bert Ely*

Department of Biological Sciences, University of South Carolina, Columbia, SC, 29208, USA

*Author to whom correspondence should be addressed

Tel.: 803-777-2768; Fax: 803-777-4002; email: ely@sc.edu

Running head: Complex MHC evolution in *Morone* species

Abstract

The deep evolutionary history of MHC allelic lineages provides an opportunity to study the accumulation of mutations over a long time scale. In the present study, a sequence analysis of a fragment of an MHC class Ia gene was performed in four North American *Morone* species and 39 flanking region alleles was identified among 45 individuals. Furthermore, an analysis of the evolutionary dynamics of these alleles along with two microsatellite loci (SB83 and SB84) in this region provided evidence for gene duplications and extensive trans-species polymorphism. Although high levels of polymorphism were detected in *M. saxatilis* (striped bass), *M. americana* (white perch), and *M. mississippiensis* (yellow bass), extremely low levels of MHC diversity were detected in *M. chrysops* (white bass), suggesting the possibility of a severe population bottleneck for this species. The perfect microsatellite, SB83, showed extensive length variability among alleles. Polarity of base substitutions within repeats of SB83 was also detected. The structure of the short interrupted compound locus, SB84, also has evolved primarily by repeat duplications. Unlike SB83, however, the alleles of SB84 correlate with the microsatellite flanking region allelic lineages, indicating that SB84 has evolved more slowly than SB83.

Key words: MHC, *Morone*, microsatellite, trans-species polymorphism, gene duplication, population bottleneck

Introduction

The major histocompatibility complex (MHC) of jawed vertebrates is a multigene family involved in antigen presentation to immune system (Klein, 1986). A striking feature of MHC is

the extensive allelic polymorphism found in many of its genes that is thought to be maintained by pathogen-driven balancing selection (Hughes & Yeager, 1998). Allelic lineages of MHC can persist over long periods of evolutionary time, even across multiple speciation events, resulting in trans-species polymorphisms (Klein *et al*., 1998a). The deep evolutionary history of the MHC allelic lineages provides an opportunity to study the historical accumulation of mutations over a long time scale.

Microsatellites are short tandemly repeated sequence motifs (1-6 nucleotides) found throughout the genomes of many higher organisms (Tautz, 1993; Schlötterer, 2000; Ellegren, 2004). Due to their hypervariablity and ubiquitous occurrence, microsatellites are widely used in an impressive number of biological applications such as parentage analysis, gene mapping, and assessments of population structure (Schlötterer & Pemberton, 1994; Goldstein & Schlötterer, 1999). In recent years, much has been learned about the complex mutational process of microsatellites. DNA replication slippage has been described as the primary mechanism causing variation in the repeat number of microsatellites (Schlötterer & Tautz, 1992; Ellegren, 2000). Base substitution has played an essential role in microsatellite contractions, by breaking long repeat arrays into smaller units (Kruglyak *et al*., 1998), but few studies have elucidated the nature of base substitutions (interruptions) in microsatellite evolution (Brohede & Ellegren, 1999; Varela *et al*., 2008). Moreover, the genesis of microsatellites still remains a matter of debate (Schlötterer, 2000; Buschiazzo & Gemmell, 2006). Microsatellites are assumed to arise via the creation of a proto-microsatellite, i.e. a short intermediate stage with as few as 3 or 4 repeat units, which is thought to be the substrate for further expansion (Schlötterer, 2000; Buschiazzo & Gemmell, 2006). The mechanism underlying the genesis and length expansion of short proto-microsatellites, however, is still not fully understood. Results of early studies indicated that short

proto-microsatellites are highly stable and do not mutate by DNA replication slippage (Messier *et al*., 1996; Rose & Falush, 1998). Further studies, however, concluded that microsatellites with few repeats could gain and lose repeat units (Pupko & Graur, 1999; Zhu *et al*., 2000; Dieringer & Schlötterer, 2003).

The genus *Morone* is comprised of four species found in North America, including a pair of closely related species, *Morone americana* (Gmelin) and *Morone mississippiensis* Jordan & Eigenmann, and their relatives *Morone saxatilis* (Walbaum) and *Morone chrysops* (Rafinesque). The *Morone* species supported valuable commercial and recreational fisheries. The diversity of MHC is important in the immune diversity of populations. However, no analysis of MHC diversity has been conducted in *Morone* species. In this study, a sequence analysis of a fragment of an MHC class Ia gene was performed in the four *Morone* species. The fragment was from a genomic clone of striped bass, initially developed as a marker for population genetic analysis (Leclerc *et al*., 1996). When the sequence was subjected to a homology search against the five available ENSEMBL teleost whole genome database, hits on 3' downstream sequences of multiple MHC class Ia genes on groupX of the three-spined stickleback, *Gasterosteus aculeatus* L. genome were detected. The sequence was then blasted against the NCBI EST database, it was found that the whole fragment was transcripted, and the ESTs were from European seabass *Dicentrarchus labrax* L. tissues infected with *V. anguillarum* or Nodavius. All evidence demonstrated that the fragment was in the 3' untranslated region of an *Morone* MHC class Ia gene. Two microsatellite loci were identified in this region. One, designated SB83, is a perfect microsatellite and the other, designated SB84, is a short interrupted compound repeat (Fig. 1).

The primary objectives of this study were: 1) to analyze the allelic variation of the MHC class Ia fragment in the four *Morone* species; 2) to test whether there is trans-species evolution

among the *Morone* species; 3) to investigate patterns of mutations of the two closely linked microsatellite loci by superimposing the microsatellite sequence onto a phylogenetic tree constructed from the microsatellite flanking regions (MFR).

Materials and methods

Sample collection and DNA extraction

Striped bass, *M. saxatilis* (n=17), were collected from the Congaree River in South Carolina during the spawning season of 1994 and the Hudson River in New York in 1997. White bass, *M. chrysops* (n=10), were sampled from North Carolina in 1992 and from the Ohio River in Ohio in 2000. White perch, *M. americana* (n=13), were collected from North Carolina in 2000, from the Choptank River in Maryland and from the Congaree River in South Carolina in 1992. Yellow bass, *M. mississippiensis* (n=5) were obtained from Caddo Lake in Texas in 2008 and from the Mississippi River in Louisiana during the spring of 1991. A specimen of European Seabass, *Dicentrarchus labrax*, L. collected from Marseilles, France in 1995 was included as an outgroup in the phylogenetic analysis. Genomic DNA was isolated from fin clips by DNAzol Genomic DNA Isolation Reagent (DN 127, Molecular Research Center, Cincinnati, Ohio) after a proteinase K digest.

SB83 microsatellite analysis

Replication errors occurred during the amplification of perfect microsatellites due to Slipped-Strand Mispairing. To assist defining of the alleles, a primer pair (FT38, 5' - TGGGCCTGATTGGAATCAAAA - 3' and FT36, 5' - GATAGGTTGTATCAATGTTGC - 3') was developed to amplify a fragment containing the SB83 microsatellite and 140 bp of the flanking region (Fig. 1). Polymerase chain reaction (PCR) amplification was carried out in a 25 µL reaction containing 1.0 µL DNA, 2.5 µL 10× buffer (New England Biolabs, Beverly, MA), 1mM $MgCl_2$, 50 µM of each dNTP, 0.2 µM forward primer (FT38 with a 5' fluorescent FAM label; Integrated DNA Technologies, Coralville, IA), 0.2 µM reverse primer (FT36), and 1.0 U Taq DNA polymerase (New England Biolabs, Beverly, MA). Thermal cycling parameters were: 95 ºC for 3 min, then 35 cycles each at 94 ºC for 20 s, 54 ºC for 20 s and 72 ºC for 30s, followed by 1 cycle of final elongation at 72 ºC for 10 min. Amplified products were run on an ABI PRISM 3130 DNA analyzer with a CXR size standard (Promega, Madison, WI). Allele scoring was performed using GENEMAPPER software version 4.0 (Applied Biosystems, Foster City, CA).

Amplification, cloning and sequencing of the SB83/84 region

The entire region encompassing both microsatellite loci and their flanking regions was amplified using primers FT38N 5' - CCCAAGCTTGGGCCTGATTGGAATCAAAA - 3' and FT16N 5'- CCGGAATTCCGGCACTTCCTATACGTACATAGT - 3', with restriction-sites for *Eco*RI or *Hin*dIII (sequence underlined) near the 5'-ends (Fig. 1). The primers were designed according to a striped bass genomic clone sequence (Leclerc *et al*., 1996). PCR was carried out in 50µL reactions containing 1.0 µL DNA, 10.0 µL 5× Phusion HF buffer, 50 µM of each dNTP,

0.2 µM forward and reverse primers, and 1.0 U Phusion high-fidelity DNA polymerase (FINNZYMES, Espoo, Finland). The thermal cycling profile was 98 ºC for 30 s, then 30 cycles each at 98 ºC for 10 s, 54 ºC for 20 s and 72 ºC for 20 s, plus a final extension step at 72 ºC for 5 min. Amplification products were purified by GenElute PCR DNA Purification Kit (SIGMA, St. Louis, MO) and double digested with *Eco*RI and *Hin*dIII (New England Biolabs, Beverly, MA). The double digest product was purified by GenElute PCR DNA Purification Kit and ligated into pBluescript II SK + vector with T4 DNA ligase (New England Biolabs, Beverly, MA). Ligated product was transformed into *Escherichia coli* strain XL1-Blue (Stratagene, La Jolla, CA) and plated on LB agar containing 100 mg/L ampicillin, spread with 40 mL of 40 mg/mL X-gal and 40 mL of 100 mM IPTG. Eight to eighteen white clones were randomly picked for each sample and grown in 2 mL LB medium at 37 ºC for 24 hours. The plasmid DNA was isolated by GenElute Plasmid Mini-Prep Kit (SIGMA, St. Louis, MO), and the insert was sequenced using BigDye terminator Cycle Sequencing Kit with both M13 and M13 reverse primers on a 3730 DNA analyzer (Applied Biosystems, Foster City, CA).

Sequence analysis

Nucleotide sequences were aligned and edited using DNAStar software (DNAStar, Madison, WI). Discrepancies in the number of repeat units in the SB83 microsatellite occurred because of replication errors that occurred during amplification. Therefore, the result from the prior SB83 microsatellite analysis was considered to be the correct number of repeats since no cloning was involved in that analysis. The single base changes arising during PCR and the products produced by PCR-mediated recombination were identified by repeating the cloning and

sequencing experiments. Clones produced in each of the two independent experiments were considered to represent true alleles. The nucleotide sequences of the 39 confirmed flanking region alleles and the European seabass sequence have been deposited in GenBank (accession numbers (GQ497674- GQ497713).

Phylogenetic trees were reconstructed using the neighbor-joining method implemented in MEGA 4.0 software (Tamura *et al*., 2007). Since microsatellite loci are subject to recurrent mutation, only the flanking region sequences were used to infer phylogenetic relationships among alleles. The neighbor-joining tree was reconstructed based on Kimura's two-parameter distances and the pairwise deletion option. Bootstrap tests (2000 replicates) were performed to establish the reliability of the inferred topologies.

Results

Molecular diversity of the SB83-SB84 region

A total of 547 clones derived from the genomic DNA of 45 individuals from the four species of *Morone* were analyzed. A total of 39 microsatellite flanking region (MFR) alleles were detected. Extensive sequence diversity was detected both within and among individuals in the genera *M. saxatilis*, *M. americana,* and *M. mississippiensis*. The number of different alleles per individual ranged from 1 to 4 in *M. saxatilis* and *M. americana*, and 2 to 3 in *M. mississippiensis* suggesting that the locus might be duplicated in these species (Table I). In contrast, the sequence diversity in *M. chrysops* was comparatively low with only two MFR

alleles that differed at a single nucleotide position observed in ten individuals. Only one allele was detected in the single specimen of *D. labrax* that was analyzed to provide an outgroup. Phylogenetic analysis of the *Morone* MFR alleles indicated trans-species polymorphisms, with both allelic lineages and alleles shared between species (Fig. 2).

Microsatellite evolution

The perfect microsatellite SB83 is highly polymorphic and mutates at a faster rate than the associated MFR. SB83 showed extensive length variability both within MFR allelic lineages and among individuals with the same MFR allele (Fig. 2). The number of uninterrupted repeat units varied from 7 to 33. In addition to length variation, five distinct interrupted repeats resulting from six different base substitutions were detected in 10 individuals among the *Morone* species sampled (Fig. 2). A preference for base substitution of G residues compared to T residues in the repeat was observed. Of the six base substitutions observed, five involved substitutions of a G and only one involved a substitution of a T (Fig. 2).

Sequencing analysis revealed 15 alleles of the SB84 microsatellite that included four variable parts: (GT)x, GCYY, (GT)y, and (AT)z. The number of repeat units in the (GT)x component ranged from 3 to 6. An (AT→GT) mutation at the 5' flanking region resulted in the addition of one repeat unit to the (GT)5 and (GT)6 alleles. The GCYY segment included two different transitions in the last two nucleotides of the GCYY sequence resulting in 3 different 4-base sequences. In addition, the GCYY segment was absent from the MFR MameMmis03 allele. The number of uninterrupted repeat units in the (GT)y component of SB84 varied from 0 to 9.

Interruptions resulting from base substitutions in (GT)y were detected in three SB84 alleles. Similarly, the terminal (AT)z sequence contained 0 to 3 repeats. Three alleles of SB84 lacked both the (GT)y and (AT)z segments entirely. Unlike SB83, the alleles of SB84 correlate with the MFR allelic lineages (Fig. 2). Thus, SB84 has a slower mutation rate than the perfect SB83 repeat. The exception to the MFR lineage correlation was that three alleles of SB84 were found in more than one MFR allelic lineage, perhaps due to recombination between alleles. The number of alleles of SB84 was high in *M. saxatilis* (7), *M. americana* (10), and *M. mississippiensis* (6). However, only one SB84 allele was detected in 10 *M. chrysops* individuals (Fig. 2).

Discussion

The phylogenetic analysis of the SB83 and SB84 flanking region sequences revealed extensive trans-species polymorphisms, which is a typical feature of MHC genes (Klein *et al*., 1998a). Trans-species persistence of allelic lineages at the MHC loci has also been documented in other teleost fishes, such as cichlids (Figueroa *et al*., 2000), cyprinids (Graser *et al*., 1996; Ottová *et al*., 2005), and salmonids (Aguilar & Garza, 2007). In addition, gene duplication has played a crucial role during the evolution of the MHC gene family. The MHC genes have been subjected to repeated cycles of expansions and contractions during their evolution, and as a consequence, a given region may harbor multiple highly related genes (Klein *et al*., 1993). Furthermore, the number of genes in each MHC class and subclass has been shown to vary considerably from species to species, and often within species (Klein *et al*., 1998b). Genomic studies have demonstrated that the number of MHC class Ia genes varies considerably among

219 teleost fish species (Clark *et al*., 2001). For example, in East African cichlid fishes, the number
220 of MHC class II B loci per haplotype varies from individual to individual, ranging from 1 to 13
221 (Málaga-Trillo *et al*., 1998). Similarly, more than two alleles per individual were detected in
222 some individuals of *M. saxatilis*, *M. americana*, and *M. mississippiensis*, indicating that this
223 locus was duplicated either in some individuals or in all individuals in these species.
224
225 Both striped bass and white bass have low levels of polymorphisms compared to other
226 species of fish (Rogier *et al*., 1985; Leclerc *et al*., 1996; Diaz *et al*., 1998; White, 2000; Han &
227 Ely, 2002). However, the level of polymorphism is five times lower in white bass than in striped
228 bass (Han & Ely, 2002). Similar results were obtained in this study where white bass had
229 extremely lower levels of polymorphism than any of the other *Morone* species. In most species, a
230 high level of genetic diversity has been maintained at the MHC locus by balancing selection
231 (Hughes & Yeager, 1998). The lower levels of polymorphism suggest that a dramatic population
232 bottleneck may have reduced the MHC diversity in white bass. Low levels of MHC
233 polymorphism have been reported among other species known to have gone through population
234 bottlenecks (Mikko & Andersson, 1995; Mikko *et al*., 1999; Miller & Lambert, 2004; Wan *et al*.,
235 2006; Mainguy *et al*., 2007).
236
237 Identical numbers of SB83 repeats were observed in different MFR allelic lineages both
238 in the same species and in different *Morone* species (Fig. 2). Thus, contemporary SB83 allele
239 sizes cannot be used to reconstruct the evolutionary history of this complex locus. Polarity of
240 base substitutions within the repeats of SB83 also was observed. Two substitutions occurred at a
241 terminal repeat and the other three substitutions occurred in the fourth or fifth repeat from an end.

Similar results were reported in previous studies focusing on base substitution within microsatellite repeat arrays. Through sequence analysis of 22 orthologous bovine and ovine (CA)n loci, Brohede & Ellegren (1999) found that the substitution rate in the end of microsatellites was significantly higher than that in the middle of repeat regions. These authors proposed three models that could account for the relatively high mutation rate at the ends of the arrays, all of which were associated with the inefficiency of the mismatch repair system during either replication or recombination. Varela *et al*. (2008) explored the distribution of interruptions in dinucleotide repeats from the human genome and found that the interruptions tended to be towards the ends of microsatellites as well. Another feature of SB83 is that five of the six observed base substitutions occurred at a G in the SB83 microsatellite. A similar result was observed in dinucleotide repeats of human genome. Varela *et al*. (2008) found that mutations involving the substitution of C were approximately 3 times more frequent than those involving substitutions of A in (AC)10.

In contrast to SB83, SB84 is a compound system of repeats with a complex mutation pattern and a slower rate of evolution. Several alleles of SB84 were found in more than one species suggesting that the complexity at this locus predates speciation. Also, three SB84 alleles were shared by distantly-related flanking region alleles suggesting that intragenic recombination may play a role in generating diversity in the SB83/SB84 region as well.

In conclusion, high levels of allelic diversity and gene duplication were detected in the MHC class Ia fragment of *M. saxatilis*, *M. americana*, and *M. mississippiensis*. Extremely low levels of diversity were detected in *M. chrysops*, which could be the result of severe population

bottleneck. Extensive trans-species polymorphisms were demonstrated among the *Morone* species. Theses results suggested that patterns of allelic variation of the MHC class Ia fragment in the *Morone* species were regulated by balancing selection. Both repeat duplications and polarity of base substitutions were detected in the perfect microsatellite locus, SB83. However, the short interrupted compound locus, SB84, evolved primarily by repeat duplications and more slowly than SB83.

Acknowledgements

We thank Jeremy Dietrick and Kurt Ash for technical assistance, Billy Carter for providing yellow bass samples, and the two anonymous reviewers for their valuable comments. This work was supported in part by FISHTEC grant NOAA-NMFS RT/F-1.

References

Aguilar, A. & Garza, J.C. (2007) Patterns of historical balancing selection on the salmonid major histocompatibility complex class II β gene. *Journal of Molecular Evolution*, **65**, 34-43. doi: 10.1007/s00239-006-0222-8

Brohede, J. & Ellegren, H. (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proceedings of the Royal Society of London. Serials B: Biological Sciences* **266**, 825-833.

Buschiazzo, E. & Gemmell, N.J. (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **28**, 1040-1050. doi: 10.1002/bies.20470

287 Clark, M.S., Shaw, L., Kelly, A., Snell, P. & Elger, G. (2001) Characterization of the MHC class
288 I region of the Japanese pufferfish (*Fugu rubripes*). *Immunogenetics* **52**, 174-185. doi:
289 10.1007/s002510000285
290 Diaz, M., Macpherson, J. & Ely, B. (1998) Striped bass population subdivision within the
291 Santee-Cooper system, South Carolina. *Molecular Marrine Biology and Biotechnology*
292 **7**(3), 191-196.
293 Dieringer, D. & Schlötterer, C. (2003) Two distinct modes of microsatellite mutation processes:
294 evidence from the complete genomic sequences of nine species. *Genome Research* **13**,
295 2242-2251. doi: 10.1101/gr.1416703
296 Ellegren, H. (2000) Microsatellite mutations in the germline: implications for evolutionary
297 inference. *Trends in Genetics* **16**, 551-558. doi: 10.1016/S0168-9525(00)02139-9
298 Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Review*
299 *Genetics* **5**, 435-445. doi: 10.1038/nrg1348
300 Figueroa, F., Mayer, W.E., Sültmann, H., O'hUigin, C., Tichy, H., Satta, Y.,Takezaki, N.,
301 Takahata, N. & Klein, J. (2000) Mhc class II *B* gene evolution in East African cichlid
302 fishes. *Immunogenetics* **51**, 556-575. doi: 10.1007/s002510000181
303 Goldstein, D.B. & Schlötterer, C. (1999) Microsatellites: Evolution and Applications. Oxford
304 University Press, USA.
305 Graser, R., O'hUigin, C., Vincek, V., Meyer, A. & Klein, J (1996) Trans-species polymorphism
306 of class II *Mhc* loci in danio fishes. *Immunogenetics* **44**, 36-48. doi: 10.1007/BF02602655
307 Han, K. & Ely, B. (2002) Use of AFLP analyses to assess genetic variation in *Morone* and
308 *Thunnus* species. *Marine Biotechnology* **4**, 141-145. doi: 10.1007/s10126-001-0080-8

Hughes, A.L. & Yeager, M. (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32**, 415-435. doi: 10.1146/annurev.genet.32.1.415

Klein, J. (1986) Natural history of the Major Histocompatibility Complex. New York: Wiley.

Klein, J., Ono, H., Klein, D. & O'hUigin, C. (1993) The accordion model of Mhc evolution. In: Gergely, J., Petranyi, G. (Eds), Progress in Immunology. Heidelberg: Springer-Verlag, pp. 137-143

Klein, J., Sato, A., Nagl, S. & O'hUigin, C. (1998a) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics* **29**, 1-21. doi: 10.1146/annurev.ecolsys.29.1.1

Klein, J., Sato, A. & O'hUigin, C. (1998b) Evolution by gene duplication in the major histocompatibility complex. *Cytogenetics and Cell Genetics* **80**, 123-127. doi: 10.1159/000014967

Kruglyak, S., Durrett, R.T., Schug, M.D. & Aquadro, C.F. (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10774-10778.

Leclerc, G.M., Diaz, M. & Ely, B. (1996) Use of PCR-RFLP assays to detect genetic variation at single copy nuclear loci in striped bass (*Morone saxatilis*). *Molecular Marine Biology and Biotechnology* **5**(2), 138-144.

Mainguy, J., Worley, K., Côté SD. & Coltman, D.W. (2007) Low MHC DRB class II diversity in the mountain goat: past bottlenecks and possible role of pathogens and parasites. *Conservation Genetics* **8**, 885-891. doi: 10.1007/s10592-006-9243-5

Málaga-Trillo, E., Zaleska-Rutczynska, Z., McAndrew, B., Vincek, V., Figueroa, F., Sültmann, H. & Klein, J. (1998) Linkage relationships and haplotype polymorphism among cichlid MHC class II B loci. *Genetics* **149**, 1527-1537.

Messier, W., Li, S.H. & Stewart, C.B. 1996. The birth of microsatellites. *Nature* **381**, 483.

Mikko, S. & Andersson, L. (1995) Low major histocompatibility complex class II diversity in European and North American moose. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 4259-4263.

Mikko, S., Røed, K., Schmutz, S. & Andersson, L. (1999) Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. *Immunological Reviews* **167**, 169-178. doi: 10.1111/j.1600-065X.1999.tb01390.x

Miller, H.C. & Lambert, D.M. (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology* **13**, 3709-3721. doi: 10.1111/j.1365-294X.2004.02368.x

Ottová, E., Šimková, A., Martin, J-F., Bellocq, J.G.de, Gelnar, M., Allienne, J.F. & Morand, S. (2005) Evolution and trans-species polymorphism of MHC class IIβ genes in cyprinid fish. *Fish & Shellfish Immunology* **18**, 199-222. doi: 10.1016/j.fsi.2004.07.004

Pupko, T. & Graur, D. (1999) Evolution of microsatellites in the yeast *Sacharomyces cerevisiae*: role of length and number of repeated units. *Journal of Molecular Evolution* **48**, 313-316. doi: 10.1007/PL00006474

Rogier, C.G., Ney, J.J. & Turner, B.J. (1985) Electrophoretic analysis of genetic variability in a landlocked striped bass population. *Transactions of the American Fisheries Society* **114**, 244-249.

353 Rose, O. & Falush, D. (1998) A threshold size for microsatellite expansion. *Molecular Biology*
354 *and Evolution* **15**,613-615.
355 Schlötterer, C. & Tautz, D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids*
356 *Research* **20**, 211-215.
357 Schlötterer, C. & Pemberton, J. (1994) The use of microsatellites for genetic analysis of natural
358 populations. In: Schierwater, B., Streit, B., Wagner, G., DeSalle, R., (Eds), Molecular
359 ecology and evolution: approaches and applications. Birkhäuser Verlag, Basel,
360 Switzerland, pp. 203-214.
361 Schlötterer, C. (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371.
362 doi: 10.1007/s004120000089
363 Tautz, D. (1993) Notes on the definition and nomenclature of tandemly repetitive DNA
364 sequences. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T., Jeffreys, A.J., (Eds), DNA
365 fingerprinting: State of the Science. Birkhäuser Verlag, Basel, Switzerland, pp. 21-28.
366 Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetic
367 Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.
368 Varela, M.A., Sanmiguel, R., Gonzalez-Tizon, A. & Martinez-Lage, A. (2008) Heterogeneous
369 nature and distribution of interruptions in dinucleotides may indicate the existence of
370 biased substitutions underlying microsatellite evolution. *Journal of Molecular Evolution*
371 **66**, 575-580. doi: 10.1007/s00239-008-9107-3
372 Wan, Q.H., Zhu, L., Wu, H. & Fang, S.G. (2006) Major histocompatibility complex class II
373 variation in the giant panda (*Ailuropoda melanoleuca*). *Molecular Ecology* **15**, 2441-
374 2450. doi: 10.1111/j.1365-294X.2006.02966.x

White, M.M. (2000) Genetic variation in white bass. *Transactions of the American Fisheries Society* **129**, 879-885.

Zhu, Y., Strassmann, J.E. & Queller, D.C. (2000) Insertions, substitutions, and the origin of microsatellites. *Genetical Research* **76**, 227-236.

Table I, List of samples analyzed in the present study. The number of clones sequenced, alleles, and name of Microsatellite flanking region (MFR) alleles were also shown. The MFR sequence designations were defined by their organism origins. Species are designated by four-letter abbreviations of the Latin names (one from the first letter of genus name and three from the first three letters of species name, e.g. *Msax* for *Morone saxatilis*).

Specimen	Sample location	Number clones sequenced	Number of alleles	Name of MFR alleles
M. mississippiensis-01	Mississippi River, LA	12	3	*Mmis04, Mmis01, MsaxMmis01*
M. mississippiensis-02	Mississippi River, LA	8	2	*Mmis04*
M. mississippiensis-03	Caddo Lake, TX	12	3	*Mmis03, MameMmis01, MameMmis03*
M. mississippiensis-04	Caddo Lake, TX	10	3	*Mmis02 MameMmis01, MchrMmis01,*
M. mississippiensis-05	Caddo Lake, TX	12	3	*MameMmis01, MameMmis02, MchrMmis01,*
M. chrysops-01	North Carolina	10	1	*Mchr01*
M. chrysops-02	North Carolina	8	1	*Mchr01*
M. chrysops-03	North Carolina	8	2	*MchrMmis01, Mchr01*
M. chrysops-04	North Carolina	8	1	*MchrMmis01*
M. chrysops-05	North Carolina	8	1	*Mchr01*
M. chrysops-06	North Carolina	8	2	*Mchr01*
M. chrysops-07	North Carolina	8	1	*MchrMmis01*
M. chrysops-08	Ohio River, OH	8	2	*Mchr01*
M. chrysops-09	Ohio River, OH	8	1	*Mchr01*
M. chrysops-10	Ohio River, OH	8	2	*MchrMmis01*
M. americana-01	Congaree River, SC	12	2	*Mame1, Mame16*
M. americana-02	Congaree River, SC	10	2	*MameMmis01, Mame11*
M. americana-03	Congaree River, SC	8	1	*Mame06*

M. americana-04	Congaree River, SC	12	3	*Mame07, Mame08, MameMmis03*
M. americana-05	Congaree River, SC	12	4	*MameMmis01, Mame4, Mame5, Mame9*
M. americana-06	North Carolina	12	3	*Mame01, Mame16*
M. americana-07	North Carolina	10	3	*Mame02, Mame12, Mame15*
M. americana-08	North Carolina	12	2	*Mame03, MameMmis03*
M. americana-09	Choptank River, MD	14	4	*Mame10, Mame16, MameMmis03*
M. americana-10	Choptank River, MD	14	4	*MameMmis02, Mame13, Mame14, MameMmis03*
M. americana-11	Choptank River, MD	12	3	*MameMmis02, Mame16, MameMmis03*
M. americana-12	Choptank River, MD	8	4	*MameMmis02, MameMmis03*
M. americana-13	Choptank River, MD	10	3	*Mame02, Mame12, Mame15*
M. saxatilis-01	Congaree River, SC	24	3	*Msax08, Msax13*
M. saxatilis-02	Congaree River, SC	22	3	*Msax04, Msax07, Msax09*
M. saxatilis-03	Congaree River, SC	24	2	*Msax04, Msax13*
M. saxatilis-04	Congaree River, SC	12	2	*Msax04, MsaxMmis01*
M. saxatilis-05	Congaree River, SC	23	2	*Msax12, Msax13*
M. saxatilis-06	Congaree River, SC	12	2	*Msax04, Msax13*
M. saxatilis-07	Congaree River, SC	12	2	*Msax09, Msax11*
M. saxatilis-08	Congaree River, SC	12	2	*Msax09, Msax11*
M. saxatilis-09	Congaree River, SC	8	2	*Msax02, Msax03*
M. saxatilis-10	Hudson River, NY	12	2	*Msax08, MsaxMmis01*
M. saxatilis-11	Hudson River, NY	12	3	*Msax07, Msax08, MsaxMmis01*
M. saxatilis-12	Hudson River, NY	12	3	*Msax05, Msax09*
M. saxatilis-13	Hudson River, NY	12	2	*Msax09, Msax10*
M. saxatilis-14	Hudson River, NY	12	1	*Msax01*
M. saxatilis-15	Hudson River, NY	18	4	*Msax05, Msax07, Msax09*
M. saxatilis-16	Hudson River, NY	16	3	*Msax06, Msax09, Msax10*
M. saxatilis-17	Hudson River, NY	12	2	*Msax04, Msax09*
D. labrax-01	Marseilles, France	10	1	*Dlab01*
		Total. 547		

387

Figures

Figure 1. Schematic of the two microsatellites and their flanking regions. The two sets of primers, FT38N/FT16N and FT38/FT36, amplify the whole region and the SB83 region, respectively.

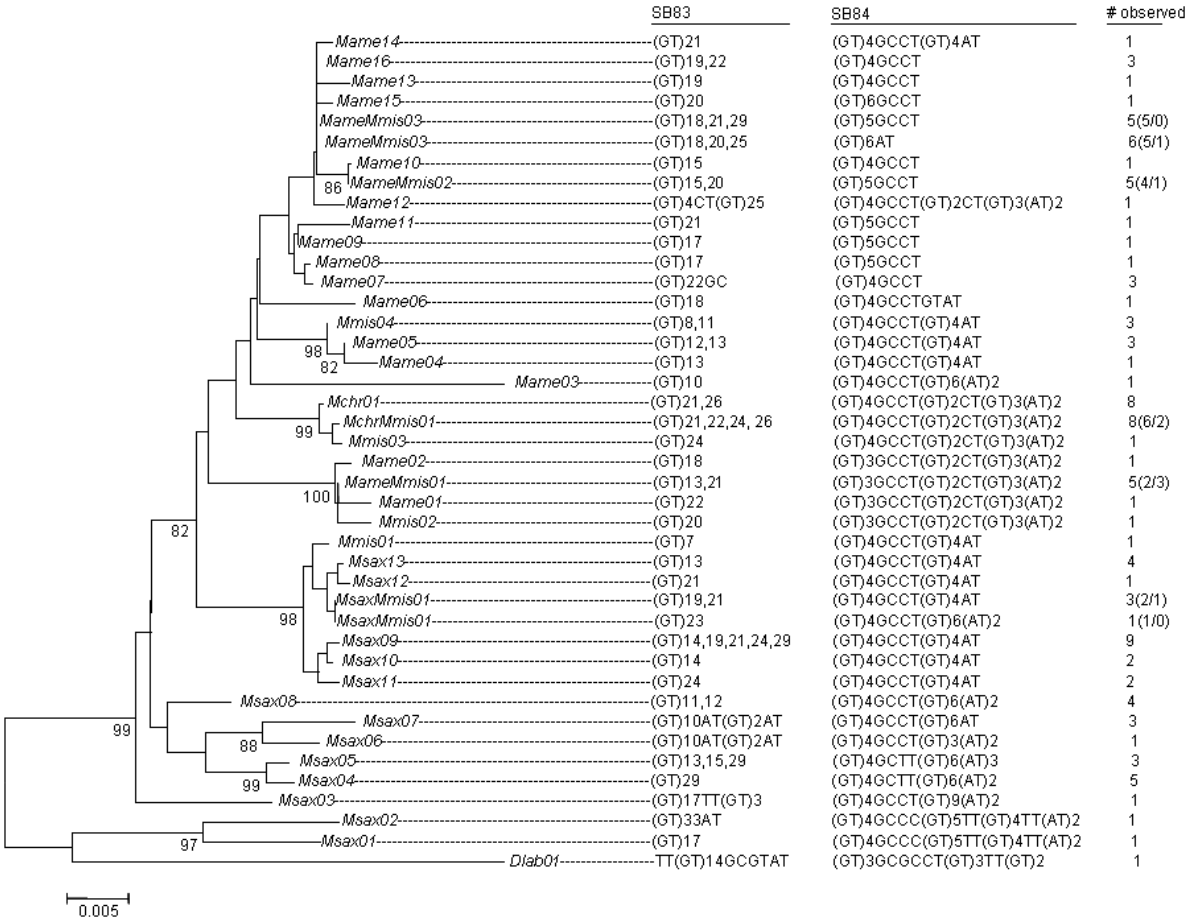

Figure 2. Neighbor-joining tree based on 39 MFR sequences with both microsatellite sequences superimposed. Observed instances of each flanking region sequence are shown. Nomenclature of the flanking region alleles were designated as in Table I. Numbers in brackets indicates observed numbers of MFR sequence in each species that shared that sequence.